AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Clinical Microbiology®

MYCOBACTERIOLOGY AND AEROBIC ACTINOMYCETES

The TB Portals: an Open-Access, Web-Based Platform for Global Drug-Resistant-Tuberculosis Data Sharing and Analysis

Alex Rosenthal,[a] Andrei Gabrielian,[a] Eric Engle,[a] Darrell E. Hurt,[a] Sofia Alexandru,[f] Valeriu Crudu,[f] Eugene Sergueev,[b] Valery Kirichenko,[b] Vladzimir Lapitskii,[b] Eduard Snezhko,[b] Vassili Kovalev,[b] Andrei Astrovko,[c] Alena Skrahina,[c] Jessica Taaffe,[a] Michael Harris,[a] Alyssa Long,[a] Kurt Wollenberg,[a] Irada Akhundova,[d] Sharafat Ismayilova,[d] Aliaksandr Skrahin,[h] Elcan Mammadbayov,[d] Hagigat Gadirova,[d] Rafik Abuzarov,[d] Mehriban Seyfaddinova,[d] Zaza Avaliani,[e] Irina Strambu,[g] Dragos Zaharia,[g] Alexandru Muntean,[g] Eugenia Ghita,[g] Miron Bogdan,[g] Roxana Mindru,[g] Victor Spinu,[g] Alexandra Sora,[g] Catalina Ene,[g] Sergo Vashakidze,[e] Natalia Shubladze,[e] Ucha Nanava,[e] Alexander Tuzikov,[b] Michael Tartakovsky[a]

Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA[a]; United Institute of Informatics Problems, National Academy of Sciences of Belarus, Minsk, Republic of Belarus[b]; Republican Scientific and Practical Centre of Pulmonology and Tuberculosis, Ministry of Health, Minsk, Republic of Belarus[c]; Scientific Research Institute of Lung Diseases, Ministry of Health, Baku, Republic of Azerbaijan[d]; The National Center for Tuberculosis and Lung Diseases, Tbilisi, Republic of Georgia[e]; Phthysiopneumology Institute, Ministry of Health, Chisinau, Republic of Moldova[f]; Marius Nasta Pneumophtisiology Institute, Ministry of Health, Bucharest, Romania[g]; Belarusian State Medical University, Minsk, Republic of Belarus[h]

ABSTRACT The TB Portals program is an international consortium of physicians, radiologists, and microbiologists from countries with a heavy burden of drug-resistant tuberculosis working with data scientists and information technology professionals. Together, we have built the TB Portals, a repository of socioeconomic/geographic, clinical, laboratory, radiological, and genomic data from patient cases of drug-resistant tuberculosis backed by shareable, physical samples. Currently, there are 1,299 total cases from five country sites (Azerbaijan, Belarus, Moldova, Georgia, and Romania), 976 (75.1%) of which are multidrug or extensively drug resistant and 38.2%, 51.9%, and 36.3% of which contain X-ray, computed tomography (CT) scan, and genomic data, respectively. The top *Mycobacterium tuberculosis* lineages represented among collected samples are Beijing, T1, and H3, and single nucleotide polymorphisms (SNPs) that confer resistance to isoniazid, rifampin, ofloxacin, and moxifloxacin occur the most frequently. These data and samples have promoted drug discovery efforts and research into genomics and quantitative image analysis to improve diagnostics while also serving as a valuable resource for researchers and clinical providers. The TB Portals database and associated projects are continually growing, and we invite new partners and collaborations to our initiative. The TB Portals data and their associated analytical and statistical tools are freely available at https://tbportals.niaid.nih.gov/.

KEYWORDS tuberculosis, digital health, interactive portals, MDR-TB, *Mycobacterium tuberculosis*, query, XDR-TB, drug-resistant TB

Tuberculosis (TB) continues to represent a major health problem worldwide. An estimated one-third of the world's population is living with latent TB (1). In 2015, there were an estimated 10.4 million new (incident) TB cases worldwide, of which 5.9 million (56%) were among men, 3.5 million (34%) were among women, and 1.0 million (10%) were among children. People living with HIV accounted for 1.2 million (11%) of

Received 7 July 2017 **Returned for modification** 5 August 2017 **Accepted** 1 September 2017

Accepted manuscript posted online 13 September 2017

Citation Rosenthal A, Gabrielian A, Engle E, Hurt DE, Alexandru S, Crudu V, Sergueev E, Kirichenko V, Lapitskii V, Snezhko E, Kovalev V, Astrovko A, Skrahina A, Taaffe J, Harris M, Long A, Wollenberg K, Akhundova I, Ismayilova S, Skrahin A, Mammadbayov E, Gadirova H, Abuzarov R, Seyfaddinova M, Avaliani Z, Strambu I, Zaharia D, Muntean A, Ghita E, Bogdan M, Mindru R, Spinu V, Sora A, Ene C, Vashakidze S, Shubladze N, Nanava U, Tuzikov A, Tartakovsky M. 2017. The TB Portals: an open-access, Web-based platform for global drug-resistant-tuberculosis data sharing and analysis. J Clin Microbiol 55:3267–3282. https://doi.org/10.1128/JCM.01013-17.

Editor Geoffrey A. Land, Carter BloodCare & Baylor University Medical Center

Address correspondence to Alex Rosenthal, alexr@niaid.nih.gov.

all new TB cases (2). As one of the world's leading threats, TB was responsible for an estimated 1.4 million deaths in 2015 (2). With slightly fewer deaths than those caused by HIV, TB is the second leading cause of death by an infectious disease. The World Health Organization (WHO) Stop TB Strategy (3) and the more recent End TB Strategy (4) have significantly decreased overall TB deaths by targeting the top TB-burdened countries. Although the TB mortality rate declined by 22% between 2000 and 2015 (2), multidrug-resistant TB (MDR-TB) is rising as an imminent global health threat. In 2015, there were an estimated 250,000 deaths from MDR-TB, including rifampin-resistant TB, and it is projected that MDR-TB and extensively drug-resistant TB (XDR-TB) could become responsible for approximately 6 to 33% and almost 10% of all TB cases, respectively, by 2040 in four high-burden countries (India, the Philippines, Russia, and South Africa) (5). Even more, a larger proportion of incident MDR-TB cases will be caused by person-to-person transmission rather than acquired resistance (5). These disconcerting projections emphasize the growing global public health problem of MDR-TB. It is critical to directly address this problem and find easier and less expensive solutions to treat and eventually eradicate drug-resistant TB (DR-TB).

Globally, only 52% of MDR-TB patients were successfully treated during 2013 (6). Treatment of MDR-TB requires at least 18 months of "second-line" drugs that are generally more toxic and expensive than the standard therapeutics. Even worse, XDR-TB is treatable with even fewer drugs, including the two most effective second-line anti-TB drugs. XDR-TB occurs in about 9.5% of people with MDR-TB (7).

The complex etiology of resistant TB demands that we consider many factors to best counter the rising threat of MDR-TB. Data collection and sharing initiatives can facilitate such studies, and many TB-specific data resources are available. For example, several databases share *Mycobacterium tuberculosis* genome information, including drug resistance mutation variants, such as the Tuberculosis Database (http://www.tbdb.org/), Tuberculist (http://tuberculist.epfl.ch/), the Relational Sequencing TB Data Platform (ReSeqTB) (https://platform.reseqtb.org/), the Pathosystems Resource Integration Center (PATRIC) Bioinformatics Resource Center (BRC) (http://www.patricbrc.org/), the Genome-Based Mycobacterium Tuberculosis Variation (GMTV) Database (http://mtb.dobzhanskycenter.org/), and the TB Drug Resistance Mutation Database (TBDReaMDB) (https://tbdreamdb.ki.se/). *M. tuberculosis* metabolomics can be explored in the BioCyc database (https://mycobacterium.biocyc.org/). TuberQ (http://tuberq.proteinq.com.ar/bioflux/information/UsageExample.html) and the TB Drug Target Database (https://www.bioinformatics.org/tbdtdb/) provide information on drug targets through structural data and proteomics. FIND maintains a virtual TB strain bank with samples available through a request for research and subsequent data sharing (https://www.finddx.org/specimen-banks/), and those researchers are developing a linked TB biomarker database (8). The TB Platform for Aggregation Clinical TB Studies (TB-PACTS) (https://c-path.org/programs/tb-pacts/) makes standardized clinical trial data available to researchers.

Despite the substantial number of TB data resources, most of them specialize in a narrow type of data. Linked patient case information or samples are often also absent from them. Under the leadership of the National Institute of Allergy and Infectious Diseases (NIAID), we developed the TB Portals program, the mission of which is to unite radiological, genomic, clinical, laboratory, and socioeconomic/geographic data from prospective and retrospective TB cases and their associated clinical samples and to freely share these curated data, including powerful and user-friendly analytical tools, with researchers and health care specialists throughout the world. We accomplish this mission through (i) international collaboration with clinical research sites and academic research organizations in countries with a heavy burden of drug-resistant TB and (ii) the creation, support, expansion, and promotion of the multifactor repository of anonymized clinical data from patients with drug-resistant forms of tuberculosis. The repository with the associated country-specific portals and analytical tools is collectively known as the TB Portals.

We designed the TB Portals to harmonize data from any international collaborators

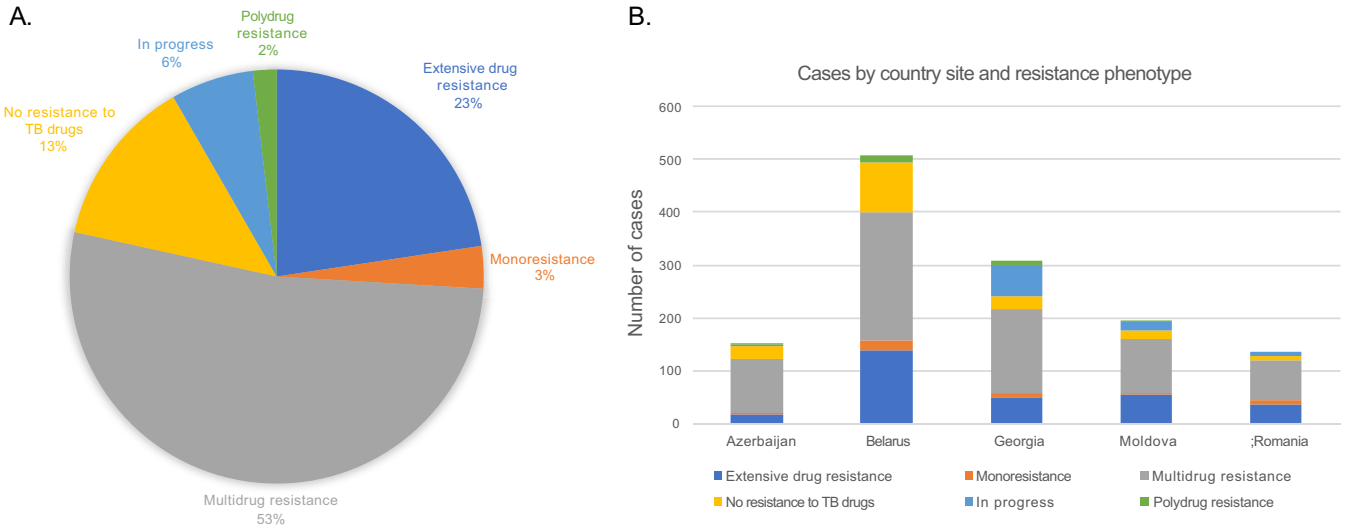

FIG 1 Drug-resistant TB cases in the TB Portals. Shown are the percentages of cases with various drug resistance phenotypes among all cases (A) and by country site (B). Resistance definitions are as follows: monoresistance is resistance to one first-line anti-TB drug only, polydrug resistance is resistance to more than one first-line anti-TB drug (other than both isoniazid and rifampin), multidrug resistance is resistance to at least both isoniazid and rifampin, and extensive drug resistance is resistance to any fluoroquinolone and to at least one of 3 second-line injectable drugs (capreomycin, kanamycin, and amikacin) in addition to multidrug resistance. "In progress" describes cases whose full case data entry is not yet complete.

and openly share them. The Web interface of the TB Portals invites users (researchers, doctors, students, biostatisticians, and policymakers) to browse and search through clinical cases of patients with MDR-TB, including all of their associated data. More importantly, through the TB Portals analytical tools, users can form virtual cohorts through complex queries of these data and run statistical analyses and visualizations of these cohorts by using a friendly graphical interface. The TB Portals allow specialists to learn from experiences of countries already under a heavy burden of MDR-TB. The cases described in the TB Portals represent real patients and are backed by physical samples that can be shared. We hope that the open data and access to the underlying samples will promote technologically advanced research projects that explore various forms of the pathogen and the disease. In particular, we encourage detailed studies on those cases that are unique or difficult to detect by using common diagnostics so as to develop recommendations for successful treatment with increased survival and shorter treatment durations for MDR-TB patients.

RESULTS

A main product of the TB Portals program is the wide array of unique and valuable data collected from TB patients and stored in a user-friendly, open-access, and easily searched and analyzed online platform. Additionally, several applied research projects (ARPs) have stemmed from the TB Portals program, including genomic sequencing, quantitative image analysis, and drug discovery efforts. Data from ARPs have subsequently been added back into the TB Portals database.

TB patient clinical data in the TB Portals. The TB Portals program has to date collected clinical, socioeconomic, genomic, and imaging data from 1,299 TB patient cases. Over three-quarters of them are cases of multidrug-resistant (52.5%) or extensively drug-resistant (22.6%) TB (Fig. 1A and Table 1). The majority of cases have been provided by the Belarus country site ($n = 507$), followed by Georgia ($n = 308$), which is indicative of the history of country site participation in the program (Table 1). Proportions of drug-resistant case profiles across country sites mimic the overall distribution of cases in the TB Portals (Fig. 1B). It is important to note that the ratio of drug-sensitive TB and MDR-TB cases in the TB Portals do not reflect their ratio in the general population. This is by design, as we are not looking for representative cohorts for epidemiological studies or drug testing. Instead, we encourage teams to collect data

TABLE 1 Summary of general clinical data, demographic data, and types of data associated with cases in the TB Portals[a]

Parameter	No. (%) of cases						Total no. of cases
	XDR	Monoresistance	MDR	NR	In progress	PDR	
Total cases[b]	294 (22.6)	43 (3.3)	682 (52.5)	172 (13.2)	84 (6.5)	24 (1.8)	1,299
X-ray data exist	113 (38.4)	18 (41.9)	234 (34.3)	77 (44.8)	40 (47.6)	14 (58.3)	496
CT data exist	170 (57.8)	26 (60.5)	341 (50.0)	116 (67.4)	9 (10.7)	13 (54.2)	675
Genomic data exist	97 (33.0)	14 (32.6)	278 (40.8)	62 (36.0)	13 (15.5)	7 (29.2)	471
Outcome							
Still on treatment	163 (55.4)	12 (27.9)	320 (46.9)	5 (2.9)	66 (78.6)	8 (33.3)	574
Completed	11 (3.7)	4 (9.3)	51 (7.5)	36 (20.9)	1 (1.2)	2 (8.3)	105
Cured	37 (12.6)	20 (46.5)	141 (20.7)	92 (53.5)	4 (4.8)	6 (25.0)	300
Default	21 (7.1)	0 (0.0)	69 (10.1)	11 (6.4)	5 (6.0)	6 (25.0)	112
Died	32 (10.9)	2 (4.7)	41 (6.0)	8 (4.7)	1 (1.2)	0 (0.0)	84
Failure	12 (4.1)	0 (0.0)	28 (4.1)	4 (2.3)	0 (0.0)	2 (8.3)	46
Unknown	7 (2.4)	2 (4.7)	8 (1.2)	14 (8.1)	1 (1.2)	0 (0.0)	32
Case definition							
Patient has never been treated for TB or has taken anti-TB drugs for <1 mo	67 (22.8)	28 (65.1)	327 (47.9)	139 (80.8)	53 (63.1)	14 (58.3)	628
Patient has previously been treated for TB	64 (21.8)	10 (23.3)	142 (20.8)	22 (12.8)	12 (14.3)	4 (16.7)	254
Patient has previously been treated for TB and was declared lost to follow-up at the end of his/her most recent course of treatment	32 (10.9)	4 (9.3)	99 (14.5)	6 (3.5)	9 (10.7)	2 (8.3)	152
Patient has previously been treated for TB, and treatment failed at the end of his/her most recent course of treatment	119 (40.5)	0 (0.0)	95 (13.9)	2 (1.2)	5 (6.0)	3 (12.5)	224
Patient has previously been treated for TB, but outcome after his/her most recent course of treatment is unknown or undocumented	12 (4.1)	0 (0.0)	17 (2.5)	2 (1.2)	2 (2.4)	1 (4.2)	34
Gender							
Female	89 (30.3)	18 (41.9)	181 (26.5)	59 (34.3)	22 (26.2)	1 (4.2)	370
Male	205 (69.7)	25 (58.1)	501 (73.5)	113 (65.7)	62 (73.8)	23 (95.8)	929
Case origination							
Azerbaijan	16 (5.4)	4 (9.3)	102 (15.0)	25 (14.5)	2 (2.4)	1 (4.2)	150
Belarus	137 (46.6)	19 (44.2)	244 (35.8)	95 (55.2)	0 (0.0)	12 (50.0)	507
Georgia	50 (17.0)	8 (18.6)	158 (23.2)	25 (14.5)	60 (71.4)	7 (29.2)	308
Moldova	53 (18.0)	5 (11.6)	101 (14.8)	18 (10.5)	14 (16.7)	4 (16.7)	195
Romania	35 (11.9)	7 (16.3)	76 (11.1)	9 (5.2)	7 (8.3)	0 (0.0)	134
Age of onset (yr)							
<15	0 (0.0)	1 (2.3)	4 (0.6)	0 (0.0)	0 (0.0)	0 (0.0)	5
15–24	34 (11.6)	3 (7.0)	89 (13.0)	21 (12.2)	6 (7.1)	2 (8.3)	155
25–34	83 (28.2)	10 (23.3)	181 (26.5)	39 (22.7)	21 (25.0)	5 (20.8)	339
35–44	74 (25.2)	8 (18.6)	145 (21.3)	31 (18.0)	17 (20.2)	7 (29.2)	282
45–54	57 (19.4)	10 (23.3)	148 (21.7)	36 (20.9)	23 (27.4)	7 (29.2)	281
55–64	40 (13.6)	6 (14.0)	86 (12.6)	26 (15.1)	13 (15.5)	3 (12.5)	174
65+	6 (2.0)	5 (11.6)	29 (4.3)	19 (11.0)	4 (4.8)	0 (0.0)	63
Comorbidity(ies)							
Liver diseases and other comorbidities	0 (0.0)	1 (2.3)	5 (0.7)	0 (0.0)	0 (0.0)	0 (0.0)	6
Liver diseases	10 (3.4)	1 (2.3)	22 (3.2)	2 (1.2)	10 (11.9)	2 (8.3)	47
Other comorbidities	32 (10.9)	2 (4.7)	41 (6.0)	2 (1.2)	7 (8.3)	0 (0.0)	84
Diabetes and HIV/AIDS	5 (1.7)	0 (0.0)	8 (1.2)	1 (0.6)	1 (1.2)	0 (0.0)	15
Diabetes	3 (1.0)	1 (2.3)	15 (2.2)	3 (1.7)	7 (8.3)	1 (4.2)	30
HIV/AIDS and other comorbidities, including liver diseases	5 (1.7)	0 (0.0)	8 (1.2)	(0.0)	2 (2.4)	0 (0.0)	15
HIV/AIDS	13 (4.4)	1 (2.3)	25 (3.7)	1 (0.6)	0 (0.0)	1 (4.2)	41

[a]Resistance definitions are as follows: monoresistance is resistance to one first-line anti-TB drug only, polydrug resistance (PDR) is resistance to more than one first-line anti-TB drug (other than both isoniazid and rifampin), multidrug resistance (MDR) is resistance to at least both isoniazid and rifampin, and extensive drug resistance (XDR) is resistance to any fluoroquinolone and to at least one of 3 second-line injectable drugs (capreomycin, kanamycin, and amikacin) in addition to multidrug resistance. "In progress" describes cases whose full case data entry is not yet complete. NR, no resistance to TB drugs.
[b]Percentage of the total number of cases in the TB Portals (n = 1,299).

TABLE 2 Summary of clinical and pathological characteristics associated with cases in the TB Portals[a]

Parameter	No. (%) of cases						Total no. of cases
	XDR	Monoresistance	MDR	NR	In progress	PDR	
Resistance test results determined by:							
GeneXpert	45 (15.3)	9 (20.9)	150 (22.0)	36 (20.9)	2 (2.4)	5 (20.8)	247
Hain test	57 (19.4)	4 (9.3)	103 (15.1)	11 (6.4)	1 (1.2)	0 (0.0)	176
Unspecified test	95 (32.3)	18 (41.9)	234 (34.3)	52 (30.2)	2 (2.4)	11 (45.8)	412
Lowenstein-Jensen test	165 (56.1)	10 (23.3)	270 (39.6)	43 (25.0)	10 (11.9)	5 (20.8)	503
No. of X-ray images							
At least 1	113 (38.4)	18 (41.9)	234 (34.3)	77 (44.8)	40 (47.6)	14 (58.3)	496
At least 2	56 (19.0)	10 (23.3)	121 (17.7)	55 (32.0)	0 (0.0)	6 (25.0)	248
At least 3	39 (13.3)	8 (18.6)	64 (9.4)	36 (20.9)	0 (0.0)	3 (12.5)	150
No. of CT scans							
At least 1	170 (57.8)	26 (60.5)	341 (50.0)	116 (67.4)	9 (10.7)	13 (54.2)	675
At least 2	68 (23.1)	11 (25.6)	156 (22.9)	40 (23.3)	9 (10.7)	1 (4.2)	285
At least 3	40 (13.6)	7 (16.3)	118 (17.3)	28 (16.3)	9 (10.7)	1 (4.2)	203
Lung localization(s)							
Extrapulmonary tuberculosis	3 (1.0)	1 (2.3)	6 (0.9)	0 (0.0)	1 (1.2)	0 (0.0)	11
Pulmonary tuberculosis	238 (81.0)	30 (69.8)	547 (80.2)	131 (76.2)	78 (92.9)	17 (70.8)	1,041
Pulmonary tuberculosis and extrapulmonary tuberculosis	52 (17.7)	12 (27.9)	124 (18.2)	40 (23.3)	2 (2.4)	7 (29.2)	237
Affecting pleura							
No	16 (5.4)	8 (18.6)	70 (10.3)	45 (26.2)	0 (0.0)	7 (29.2)	146
Yes	64 (21.8)	10 (23.3)	143 (21.0)	50 (29.1)	0 (0.0)	5 (20.8)	272
Lung cavity size (mm)							
<10–25	28 (9.5)	1 (2.3)	33 (4.8)	9 (5.2)	8 (9.5)	0 (0.0)	79
<10	22 (7.5)	6 (14.0)	69 (10.1)	20 (11.6)	7 (8.3)	5 (20.8)	129
>25	28 (9.5)	2 (4.7)	39 (5.7)	4 (2.3)	5 (6.0)	2 (8.3)	80
No lung cavity	28 (9.5)	11 (25.6)	130 (19.1)	63 (36.6)	9 (10.7)	6 (25.0)	247
No. of regimens during treatment							
1	219 (74.5)	26 (60.5)	491 (72.0)	132 (76.7)	43 (51.2)	17 (70.8)	928
2	21 (7.1)	5 (11.6)	49 (7.2)	3 (1.7)	3 (3.6)	3 (12.5)	84
3+	33 (11.2)	1 (2.3)	86 (12.6)	4 (2.3)	1 (1.2)	2 (8.3)	127

[a]Resistance definitions are as follows: monoresistance is resistance to one first-line anti-TB drug only, polydrug resistance is resistance to more than one first-line anti-TB drug (other than both isoniazid and rifampin), multidrug resistance is resistance to at least both isoniazid and rifampin, and extensive drug resistance is resistance to any fluoroquinolone and to at least one of 3 second-line injectable drugs (capreomycin, kanamycin, and amikacin) in addition to multidrug resistance. "In progress" describes cases whose full case data entry is not yet complete.

from the most virulent, drug-resistant, and/or deadly cases, because by studying them, we can pinpoint what distinguishes them from readily available reference strains.

Twenty-three percent of cases in the TB Portals have been cured (defined by WHO criteria [9]), and 10% of cases have failed treatment or died (Table 1). These numbers represent case outcomes and the status of cases in the TB Portals at the time of manuscript writing, as 574 cases (44.2% of total cases) are currently still in treatment. Other patient case history and demographic information, such as case definition, gender, age of onset, and comorbidities, is summarized in Table 1.

Table 2 summarizes clinical characteristics of data in the TB Portals, specifically the number and percentage of cases with radiological data, drug resistance tests, types of pathology, and multiple-drug regimens. All cases have results from at least one drug resistance test, which confirms sensitivity (Fig. 2A) or resistance (Fig. 2B) to at least 1 of 16 different TB drugs. Drug resistance tests include both molecular and culture-based methods, specifically Bactec, Hain, GeneXpert, and Lowenstein-Jensen testing.

Bacterial whole-genome sequencing and sample availability. The TB Portals Steering Committee directed the sequencing and genomic study of *M. tuberculosis* found in patients with MDR-TB and XDR-TB to enhance our understanding of the molecular basis of the disease, as genomic variations are known to be related to the resistance of *M. tuberculosis* to common antibiotics (10–16).

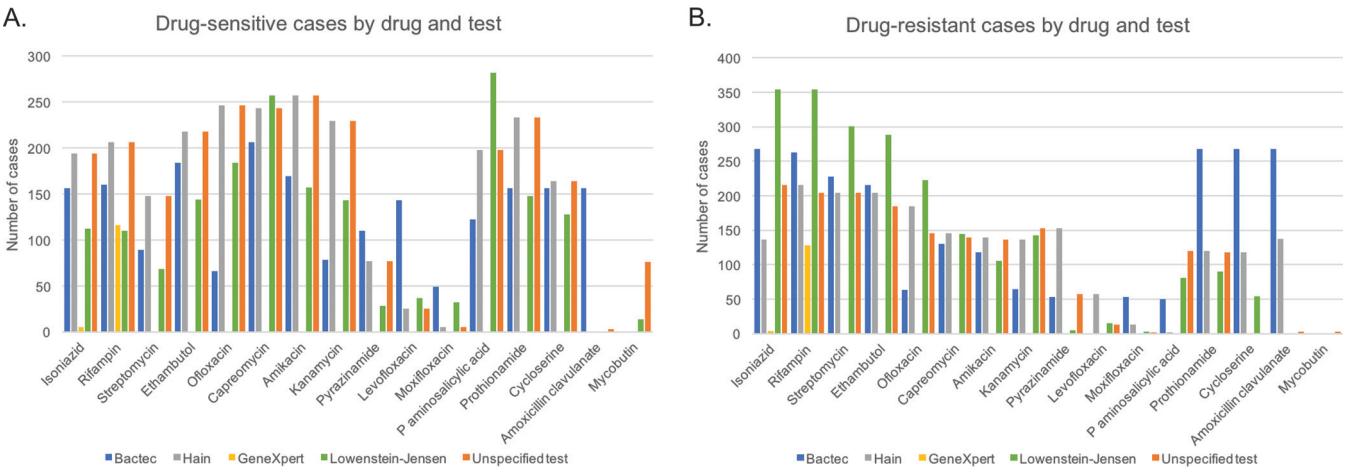

FIG 2 Resistances and sensitivities of *M. tuberculosis* samples from cases in the TB Portals to common TB drugs. Shown are the numbers of cases with samples determined to be sensitive (A) or resistant (B) to common TB drugs, as indicated by Bactec, Hain, GeneXpert, or Lowenstein-Jensen testing. "Unspecified test" describes entries for which the specific drug sensitivity test used was not entered.

To date, genomic data have been obtained for 471 (36.3%) cases (Table 1). The largest proportion of samples collected (some cases may have multiple associated samples) at the Belarus, Azerbaijan, and Georgia country sites comes from the Beijing lineage, followed by T1. Samples from the Moldova country site are almost equally split between Beijing and H3, and the T1 lineage predominates among the 49 Romanian samples (Fig. 3). Single nucleotide polymorphisms (SNPs) that target the *katG*, *rpoB*, and *gyrA* genes occur most frequently among sequenced samples from the TB Portals program, conferring resistance to isoniazid, rifampin, ofloxacin, and moxifloxacin, as described previously by Desjardins et al. (11) (Table 3).

Quantitative medical imaging analysis. Notably, 38.2% and 51.9% of cases in the TB Portals have linked X-ray and computed tomography (CT) scan data, respectively (Table 1). Due to the richness of annotated clinical image data in the TB Portals, many research projects have already used CT and chest X-ray (CXR) images from the TB Portals to develop predictive algorithms and analysis tools that are hoped to assist radiologists and clinicians to better diagnose, monitor, and treat tuberculosis (17–19). For instance, one algorithm has been proven to accurately predict the prognosis of TB patients from CT data alone (20). Another computational tool has been proven to discriminate between *M. tuberculosis* infections and other kinds of lung lesions from radiographs alone (17). When modifications of this tool were applied to data from the TB Portals in an effort to automatically predict the drug resistance of the pathogen from imaging data alone, performance reached 75% correct assignments. Performance will further improve as the database grows.

Expression of cytochromes P450 from drug-resistant strains of *M. tuberculosis*. Full-genome sequencing of multiple *M. tuberculosis* strains enabled a comparative analysis of genes and proteins in MDR-TB and XDR-TB associated with drug metabolism. Based on these results, several variant P450 proteins corresponding to resistant TB strains have been expressed and isolated for further studies of drug metabolism and conformational changes that might aid in the design of new antituberculosis drugs.

Data retrieval and analysis within the TB Portals platform. The TB Portals contain a well-populated and multidisciplinary data set, which can be easily searched and viewed by using filters that permit the selection of subsets or cohorts of patient data based on disease or patient characteristics important to the end user (Fig. 4). For example, by using these filters, one can easily identify all patient cases in a specific country where the age of onset was 45 to 64 years, the gender is male, and the patient died as a result of infection with drug-resistant TB and then view or download all of the associated demographic, social, clinical, imaging, and genomic data. In this way, users at any level can conduct extensive data searches to form custom case cohorts.

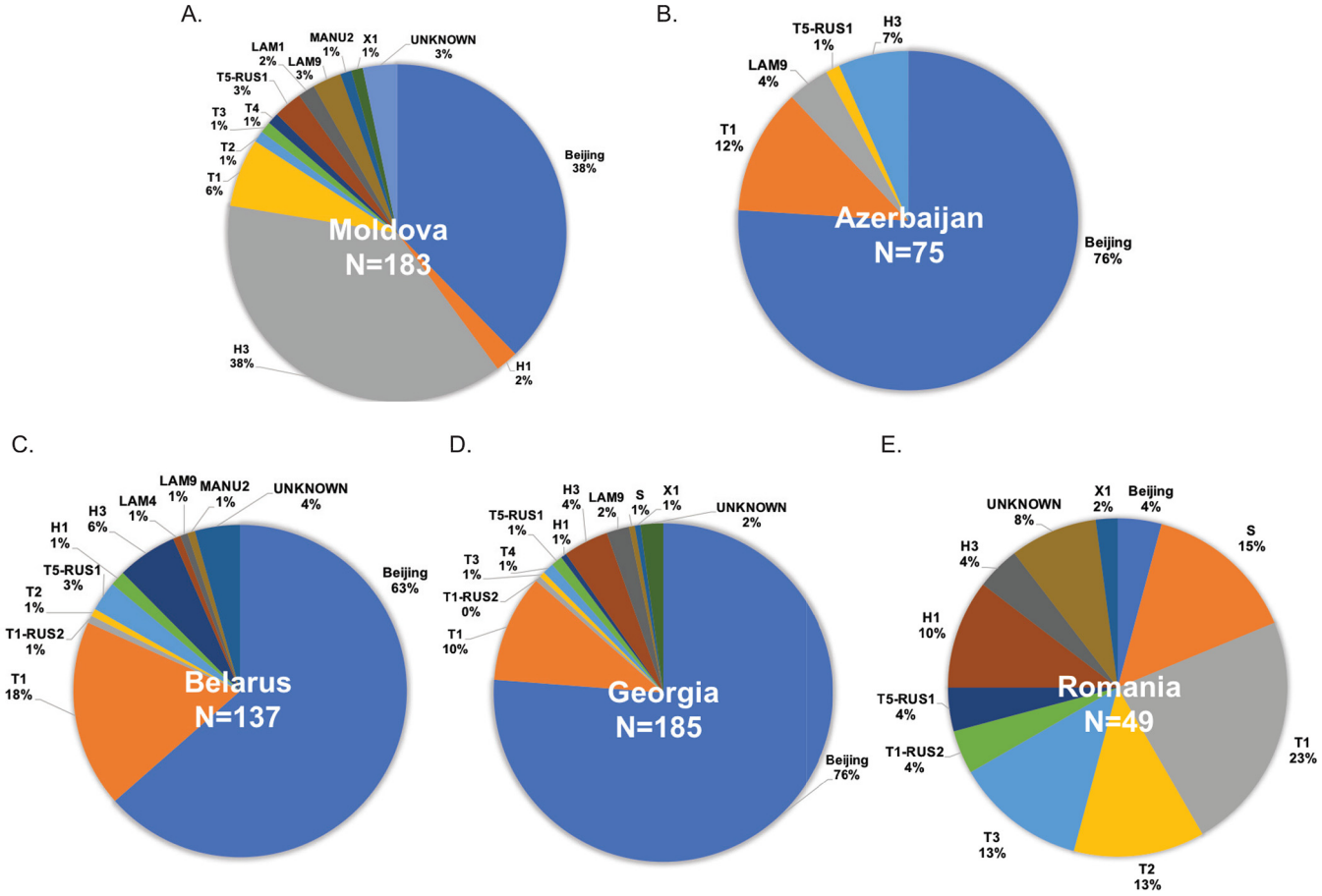

FIG 3 *M. tuberculosis* lineages among samples collected by the TB Portals program. Shown are the percentages of various *M. tuberculosis* lineages among samples originating from Moldova (A), Azerbaijan (B), Belarus (C), Georgia (D), and Romania (E) country sites.

In addition to viewing or downloading custom cohorts, physicians and health care researchers can easily compare these cohorts with statistical rigor in the TB Portals Data Exploration Portal (DEPOT) (https://depot.tbportals.niaid.nih.gov/). The DEPOT capitalizes on the uniformly annotated data in the TB Portals to enable users to query the entire database, create cohorts of patients with the same or similar clinical/socioeco-

TABLE 3 Top occurring SNPs among sequenced samples in the TB Portals program

SNP	No. of sequenced samples with SNP	% of sequenced samples with SNP among total sequenced samples	Gene affected	Drug(s) against which SNP confers resistance[a]
katG-S315T	477	75.8	*katG*	Isoniazid
rpoB-S450L	390	62.0	*rpoB*	Rifampin
gyrA-S95T	360	57.2	*gyrA*	Ofloxacin, moxifloxacin
rpsL-K43R	251	39.9	*rpsL*	Streptomycin
rrs-A1400G	129	20.5	*rrs*	Amikacin, kanamycin, capreomycin
rpsL-K88R	121	19.2	*rpsL*	Streptomycin
embB-M306V	116	18.4	*embB*	Ethambutol
embB-M306I	95	15.1	*embB*	Ethambutol
gyrA-D94G	89	14.1	*gyrA*	Ofloxacin
gyrA-A90V	69	11.0	*gyrA*	Ofloxacin
inhA-C15T	63	10.0	*inhA*	Isoniazid

[a]SNPs described by the ReSeqTB database (https://platform.reseqtb.org/) as having either "large and often conclusive evidence" or "moderate evidence for drug resistance."

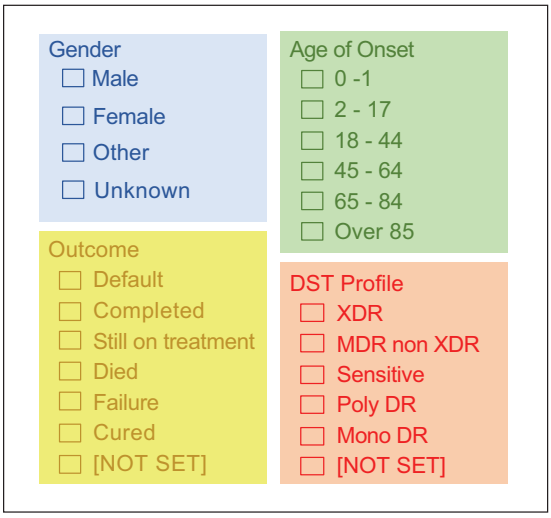

FIG 4 Queryable data. With filters, users can query data to focus on critical characteristics.

nomic/genomic descriptors, and perform comparative multifactor analyses of the most significant determinants of treatment success and survival.

DISCUSSION

The TB Portals as a multinational catalog of rare samples and unique data. Worldwide collaboration is required to reduce the threat of drug-resistant TB because it is difficult to comprehensively research all aspects of the disease without such collaboration. The TB Portals program provides a platform that enables participants to share relevant genomic, radiological, clinical, social, and research data, resulting in a large pool of fully featured drug-resistant TB cases accessible and retrievable by anyone for analysis. The data also serve as an index to samples from various countries, promoting technologically advanced research projects exploring drug-resistant *M. tuberculosis*.

Teams from the five countries (Belarus, Georgia, Azerbaijan, Moldova, and Romania), the founding participants in the program, agreed to collect, anonymize, uniformly annotate, and grant access to patient data from drug-resistant TB cases that would otherwise be unavailable to the international research community. Because the data in the TB Portals are backed by physical repositories of clinical samples in their various countries of origin, the TB Portals can also be considered an annotated catalog of TB patient samples. These samples present a unique opportunity for more targeted research, allowing the selection of specific samples of interest based on database queries and advanced data analysis capabilities provided by the TB Portals. Clinical samples, such as sputum samples, can be stored frozen for years. Cases of interest (such as when the patient was not successfully treated or had unexpected or unexplained adverse events) can be identified within the TB Portals, and their associated samples can then be retrieved for additional laboratory, genomic, or other testing to further elucidate potential causes of patient outcomes. In addition to cross-indexing the original samples, the country-specific TB Portals data also present opportunities for local staff to build analytical tools for retrospective studies in accordance with local laws and approvals.

With their rich content of highly annotated clinical data and accessible samples associated with these data, the TB Portals are an excellent resource to search for the most atypical, hard-to-explain, difficult-to-treat, virulent, and deadly variants of *M. tuberculosis* for experimental study. Because of the nature of patient case selection and data in the TB Portals, it is important to recognize that these studies, whether prospective or retrospective, will always be natural history studies and not epidemiological or clinical trial studies.

Case selection is reflective of the local research interests of participating country sites. For example, partners in Georgia are interested in cases that have been successfully cured by WHO criteria yet present later with lung pathology and require surgery. Therefore, many of the cases originating from Georgia match these characteristics and are often associated with surgical lung samples containing active mycobacteria. The Moldova team is interested in genomic factors in the pathogen related to TB reoccurrence and thus has contributed data from both reactivated and reinfected TB patient cases. The Belarus team has contributed data from cases with multiple annotated CT scans, and the Azerbaijan team has shared data from laboratory-confirmed cases of extrapulmonary TB. The TB Portals program is also focused on identifying genomic variants related to emerging drug resistance to new TB drugs and therefore actively seeks data from patient cases treated with bedaquiline and linezolid. Notably, while the TB Portals contain an assortment of interesting and atypical TB patient cases, all of their associated data are uniform and standardized, facilitating their analysis among the entire data set.

We acknowledge limitations in our data set due to gaps in patient case data and information. These gaps arise from the challenge of obtaining patient case data and information from institutions where clinical, laboratory, and imaging data are stored in separate physical locations. They are also the result of the case selection process by the country site research interests described above. Still, a substantial number of cases have genomic and/or image data (Table 1). We expect this number to increase with the expansion of the database, as data from additional cases are added and as we endeavor to provide more complete patient case data.

Ongoing efforts are focused on expanding collaborations with additional institutions and experts and the types and numbers of omics data and analytics within the TB Portals. Negotiations have been made, or are in advanced stages, for additional patient case data from sites in South Africa, China, India, Lithuania, and Kazakhstan. These additional patient cases will contribute to a more global representation of DR-TB data in the TB Portals, expanding into African and Asian regions and including countries with both high TB and high MDR-TB burdens (South Africa, China, and India), according to the WHO (21). Future data collection and research plans include genomic, proteomic, and microbiome studies of both the host and pathogen from all interested partners.

The promotion of data sharing and collaborative research is a key aim of the TB Portals program. In this regard, sample and data sharing for collaborative research projects is initiated through direct discussions with TB Portals program member country sites and partners. The program facilitates the identification of in-country research partners, namely through steering committee country representatives. Interested collaborators may directly reach out to us via the TB Central website.

TB Portals data to improve diagnostics. TB diagnosis can be challenging because of inconsistencies between testing methods. Although classic microbiological (sputum) culture testing is the gold standard of testing for active TB disease, other more rapid or readily available methods have been developed in response to the slow nature of culture testing. The WHO recommends line probe assays (LPAs) (the Genotype MTBDRplus and MTBDRsl Hain tests) and Xpert MTB/RIF tests to diagnose MDR-TB. However, these tests are not always reliable. The Xpert test has a suboptimal sensitivity for the detection of TB in sputum samples obtained from pediatric patients (22). Therefore, different detection methods, such as testing for specific antigens or the use of CXRs or CTs to detect changes in lungs, may provide reasonable alternatives.

The data in the TB Portals can contribute to efforts to improve existing TB diagnostics and to develop new diagnostic tests. The process of developing a new diagnostic test includes data collection, data analysis, and the construction of a predictive model. A new diagnostic test must contain an appropriate amount of properly annotated data to convey statistical power and demonstrate effectiveness. The data in the TB Portals may help inform this process.

TB Portals data analytics provide novel insights. A practical objective of the TB Portals program is to encourage better diagnostic and treatment methods for MDR-TB, but the data collected by the program also enable potentially more impactful clinical research. The TB Portals data and analytical tools provide opportunities to explore and examine discrepancies, such as why some standard treatments result in atypical outcomes. For example, in cases of groups of patients with drug-sensitive TB who would normally respond well to standard treatments but who nonetheless exhibit mortalities, one may be able to gain insight from their radiological, genomic, social, and clinical data, including records of previous treatments and comorbidities.

We designed the TB Portals data and query tools to be useful for many types of users with an interest in drug-resistant TB, including students, teachers, public health professionals, bioinformaticians, and physicians. The resource may be viewed as a vast natural history study catalog that can be accessed via the Internet. Unusual and atypical (presumably the most difficult for treatment) patient case histories are readily available and may be used to develop detailed adjustments of standard treatment plans. Medical images can be used as case studies of MDR-TB and XDR-TB by radiologists. The data in the TB Portals could be used to discover inconsistencies in commonly used diagnostics and variabilities in treatment efficiencies for the purposes of optimizing existing diagnostics and TB drugs used in public health programs. The connected patient, imaging, and genomic information of the TB Portals could be leveraged to compare genomic markers with drug sensitivity testing (DST) results and case outcomes. Finally, the data in the TB Portals could be used for professional training by health care professionals dealing with TB. This might be of the greatest benefit to physicians with limited experience in managing drug-resistant TB, who could use the TB Portals as a valuable and trustworthy reference of curated medical histories, clinical and imaging data, and now even genomic information. This would enable them to quickly develop successful strategies and best practices for patient care.

As types of data and, accordingly, the related descriptors within the database expand, so will insights that the TB Portals can provide. We plan to include additional laboratory data and the results of future omics research projects in the TB Portals. Some data might be original measurements (e.g., blood glucose level and hematocrit, etc.), and some might be derivative (the presence of a certain gene variant and genetic up- or downregulation, etc.). The choice of the most relevant and fundamental descriptors is important, because these descriptors ultimately determine the content of and the prospects for data mining and comparative statistical analyses that will elucidate the significant factors that determine the outcome of disease and the success of treatments.

Similarly, geographic expansion of patient cases will influence which descriptors are included in the database. These descriptors will be reflective of the participating institutions' research interests and specific to local and regional epidemic contexts. Therefore, the continued geographic growth of the TB Portals program will lead to more insights into DR-TB that are specific to and applicable in different epidemic settings.

Genomic insights. The various sequencing initiatives of the TB Portals program focus on unique samples from cases of totally drug-resistant TB, XDR-TB, and MDR-TB, the data of which have value for a wide range of international TB-related projects and resources beyond the TB Portals. Because we recognize this value, the program is dedicated to sharing genomic sequencing data and their connection to patient cases through commonly used platforms. We submit reads to the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA), and assembled and annotated *M. tuberculosis* genomes are available from the PATRIC BRC (http://www.patricbrc.org/). Our sequencing collaborators have added our data to data from a global collection of *M. tuberculosis* strains for subsequent analysis of *M. tuberculosis* evolution (16). We also collaborate to investigate novel variants responsible for drug resistance, as with the ReSeqTB initiative (https://platform.reseqtb.org/). In turn, we

distil the key results of such collaborative efforts and other research projects (23) and include them directly in the TB Portals.

In addition to contributing additional data to the TB Portals, genomic studies can provide valuable insight into country- and region-specific TB epidemics, including patterns of drug resistance. For example, a recently reported study of *M. tuberculosis* isolates from Belarus available through the TB Portals program showed that the majority of MDR-TB cases were transmitted from person to person, as opposed to being acquired *de novo*, and MDR-TB strains were likely to be of the Beijing lineage (23). An analysis of *M. tuberculosis* isolates from Moldova also shows evidence of substantial person-to-person transmission (our unpublished data). Such insights emphasize the need to focus on infection control and surveillance to prevent new cases of MDR-TB in this region.

Conclusion. By design, the TB Portals program encourages an international collaborative feedback loop of data sharing and analysis, which will build biomedical knowledge and yield actionable clinical insights. As the international data repository grows, we will refine our algorithms and introduce new tools. We hope to find novel genomic and/or imaging features that could be used for diagnostics and therapeutics or to better predict treatment outcomes.

M. tuberculosis is one of the most impactful pathogens in the world due to its extensive disease burden, and the rising incidence of drug resistance is a threat to the global community. We designed the TB Portals to facilitate the multinational and multidisciplinary collaboration required to address this growing threat. The goal of our international network is to facilitate a pathway from clinical samples to organized and curated data, including providing powerful and user-friendly analytical tools at the fingertips of the health services community. We invite countries burdened with TB to use our data infrastructure to record their cases of drug-resistant TB and to take advantage of our modern querying and analytic tools. If you are a data provider, analyst, bioinformatician, clinical specialist, or other health care professional, we invite you to join our mission to gain a better understanding of drug-resistant TB through international collaboration and to support and promote a new generation of diagnostics and treatments.

MATERIALS AND METHODS

Multinational and multidisciplinary collaborations. The following centers and institutes collaborate with the NIAID Office of Cyber Infrastructure and Computational Biology (OCICB) in the TB Portals program: the United Institute of Informatics Problems (UIIP), National Academy of Sciences of Belarus (NASB), Minsk, Republic of Belarus; the Republican Scientific and Practical Centre of Pulmonology and Tuberculosis, Ministry of Health, Minsk, Republic of Belarus; the Scientific Research Institute of Lung Diseases, Ministry of Health, Baku, Republic of Azerbaijan; The National Center for Tuberculosis and Lung Diseases, Tbilisi, Republic of Georgia; the Phthysiopneumology Institute, Ministry of Health, Chisinau, Republic of Moldova; and the Marius Nasta Pneumophtisiology Institute, Ministry of Health, Bucharest, Romania.

We formed teams of microbiologists, radiologists, and general TB researchers and physicians with practical knowledge and expertise in TB treatment and diagnostics at every collaborating medical research institute. These teams organize the efforts required to collect and annotate data for controlled submission to a centralized database. Because these teams are locally organized, they can quickly identify country-specific rules and regulations for sharing data, which are essential requirements for operating their own TB Portal. Each team independently anonymizes their patients and annotates their associated data using uniform and standardized country-specific methods. The teams also extract *M. tuberculosis* DNA from the collected clinical samples for submission to several DNA sequencing centers, mostly to the Broad Institute. We additionally formed informatics research teams focused on data collection, analytics, and genomics and radiomics.

Governance. We formed a steering committee comprised of representatives from all countries involved to advise and guide the TB Portals program (Fig. 5). The current structure of the steering committee reflects the fact that only a single institute or hospital from each country (except the Republic of Belarus) participates in the TB Portals program. The institute(s) from each country selects delegations of three representatives for each country to serve on the steering committee. Each delegation receives one vote in decisions regarding the program's activities, including requirements, regulations, projects, and collaborations. Importantly, representatives are generally frontline care providers or researchers, which ensures that the steering committee proactively generates practical and impactful ideas for program improvements. The steering committee evaluates the challenges and establishes best practices for ongoing projects, encourages and facilitates collaborative projects and new funding

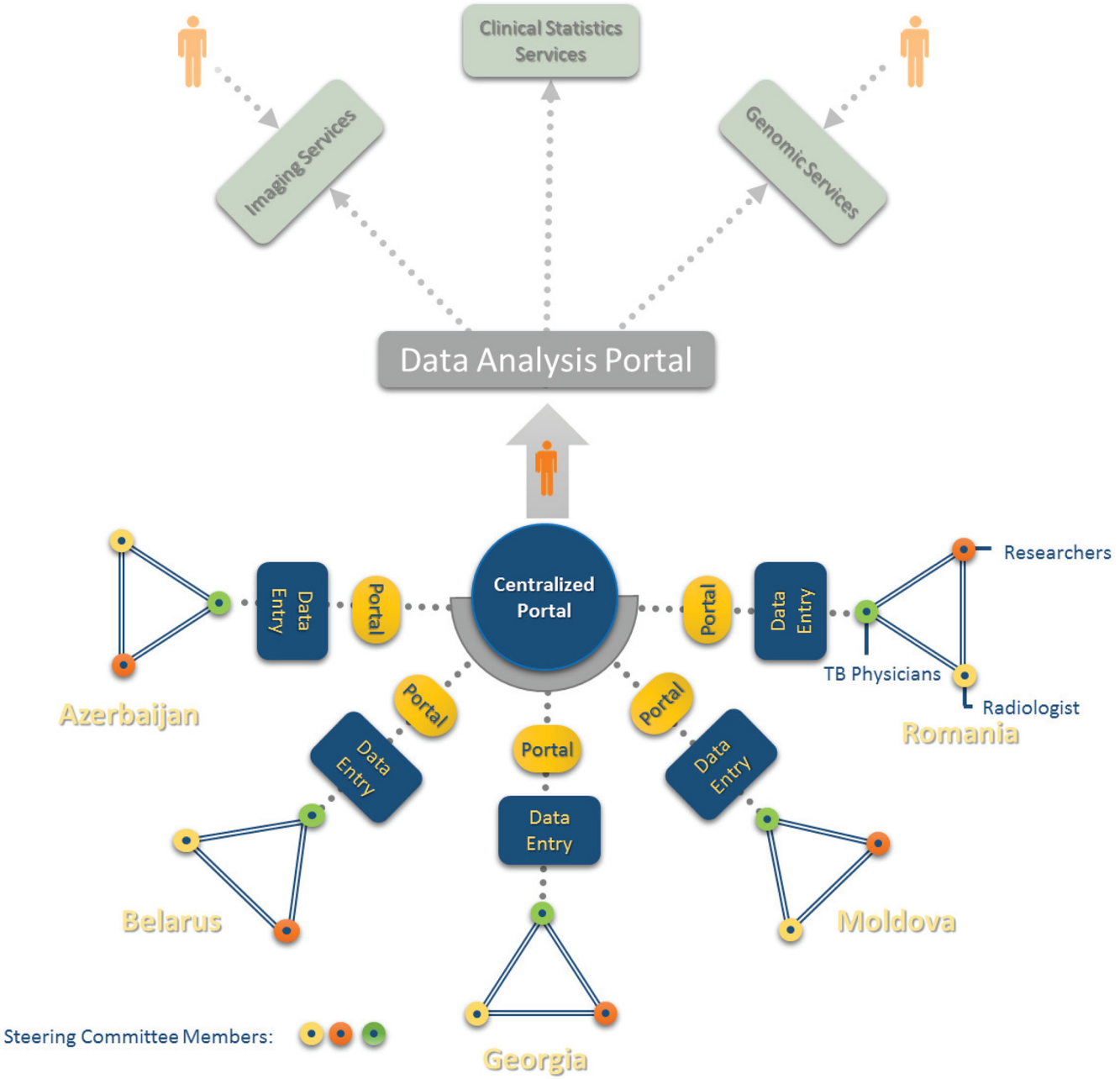

FIG 5 TB Portals data flow. Each country establishes a team. The collection of data from each team creates a centralized TB Portals site. The steering committee, including members of each country team, provides consistent guidance for the TB Portals program.

proposals, and provides feedback and guidance on the development of resources and tools for the program. Delegations and their associated teams report their current progress and deliverables during annual meetings that include workshops and presentations centered around the most important parts of the agenda.

TB Central and country-specific websites. A central TB Portals website (TB Central) (https://TBPortals.niaid.nih.gov/) was created to support the breadth of this multinational collaboration. It collects and presents information about data types, standards, analyses, participants, and publications of the program. TB Central is the hub for all TB Portals program websites, including informational pages, data sites, and analytical tools, such as the DEPOT (https://Depot.TBPortals.niaid.nih.gov/). Among the informational sites are country-specific TB Portals websites, created and maintained with user-friendly and intuitive WordPress software. Here, participants are able to present country-specific informational materials and unique research. TB Central also connects to all country-specific data sites, including those of the Republic of Azerbaijan (http://www.tuberculosis.az/), the Republic of Belarus (http://www.tuberculosis.by/), the Republic of Georgia (http://www.tuberculosis.ge/), the Republic of

TABLE 4 Description of patient data features

Class	Description of feature(s)[a]	Standard(s) used, if applicable (WHO or ICD-10, etc.)
Patient	Gender, age, ht, and wt	WHO body mass index calculated, HL7 FHIR
Social	Birthplace, habitation, history of alcohol and drug abuse, smoking, conviction, job status, history of imprisonment	No standards apply
Clinical	Symptoms, comorbidities, details of TB diagnosis (smear microscopy, rapid diagnostic test results [X-pert or Hain test], culture on liquid and solid media, DST results based on culture and nucleic acid amplification techniques, drug susceptibility/resistance profile), HIV status, cART, and records of anti-TB treatment (drugs, dosage, duration of use)	ICD-10; WHO standards; Republic Scientific and Practical Center of Pulmonology and Tuberculosis, Belarus
Imaging (X ray and CT)	Detailed annotated radiological image, including localization and size of pulmonary pathology, presence of inherited malformations and acquired complications and comorbidities, physiological respiratory status, and degree of contrast accumulation	Standard CXR
Genomic	Genomic variant data are specifically prepared for TB Portals; links to the NCBI SRA are provided when available; DNA is extracted from patient clinical samples; in some cases, genomic sequencing was retrospective (DNA from frozen and stored sputum samples was extracted for full-genome sequencing studies)	

[a]cART, combined antiretroviral therapy; ICD-10, International Classification of Diseases, volume 10; HL7, Health Level Seven International; FHIR, Fast Healthcare Interoperability Resources.

Moldova (http://www.tuberculosis.md/), and Romania (http://www.tuberculosis.ro/). TB Central Data (https://data.tbportals.niaid.nih.gov/) is the consolidated data site for all of the TB Portals program data, for users desiring to view data from all country sites.

TB Portals data entry and storage. We specifically and collaboratively developed a standardized data entry system that all participant teams uniformly use to populate the data repository (Fig. 5). Data checks and uniformity are embedded into the system using dropdown lists and ranges of permissible values. A summary of the current patient data features is shown in Table 4. The steering committee may direct the addition of new or enhanced data entry forms for the system. Because we all use the same data entry system, we can pool the data from all countries and browse, search, and analyze unique MDR-TB cases that would otherwise be impossible to share with the international scientific community.

Participant teams anonymize and deidentify individual patient information, including diagnosis, treatment history and outcome, relevant medical images (chest CT and X ray, etc.), and other types of data, in compliance with country-specific laws and regulations. Teams then manually enter the data into the TB Portals data entry site. In the future, we may consider programmatic integration with existing TB registries in current and future participant countries to enable more global data mining, but currently, all clinical and laboratory data are entered manually. All data become immediately available in both country-specific and all-country-site-consolidated direct data query tools upon publication (not entry). Publication occurs when the country-authorized physician reviewer has verified the data and determined that they are complete and accurate.

In any initiative where private patient data are involved, adherence to local laws, regulations, and requirements of clinical practice is of the utmost importance. From the very beginning of the TB Portals program, all participants agreed to be directly responsible for ensuring transparent compliance with their countries' laws, regulations, and other ethics considerations. Each country site receives approval from its clinical research facility institutional review board and follows local clinical research ethics regulations. Ethics rules required by the grant-issuing institutions (CRDF Global and the International Science and Technology Center) are strictly followed. Participant teams from each country perform quality control (QC) assessments for accuracy and the removal of personally identifiable information.

We chose the Amazon Web Services (AWS) Simple Storage Service (S3) and PostgreSQL on the AWS Relational Database Service (RDS) as the most suitable data storage resources to satisfy our requirements of functionality, extensibility, security, availability, and performance. A future report will go into further detail on the digital infrastructure behind the TB Portals.

Case data and selection criteria. All data in the TB Portals are connected to patient cases. No data are imported from elsewhere that do not belong to a patient case in the database. Each team from participating countries chooses patient cases for inclusion in their own TB Portals data repository. This selection is based on research or public health interests specific to each country site. Selection criteria also include the availability of reports on clinical outcomes and the availability of sputum samples with known drug resistance. A case may include information about additional samples collected over the course of treatment, especially if the patient continues to have positive sputum cultures. Of note, while the TB Portals contain data from drug-sensitive TB cases, we focus on the collection of data from drug-resistant, especially MDR-TB, cases.

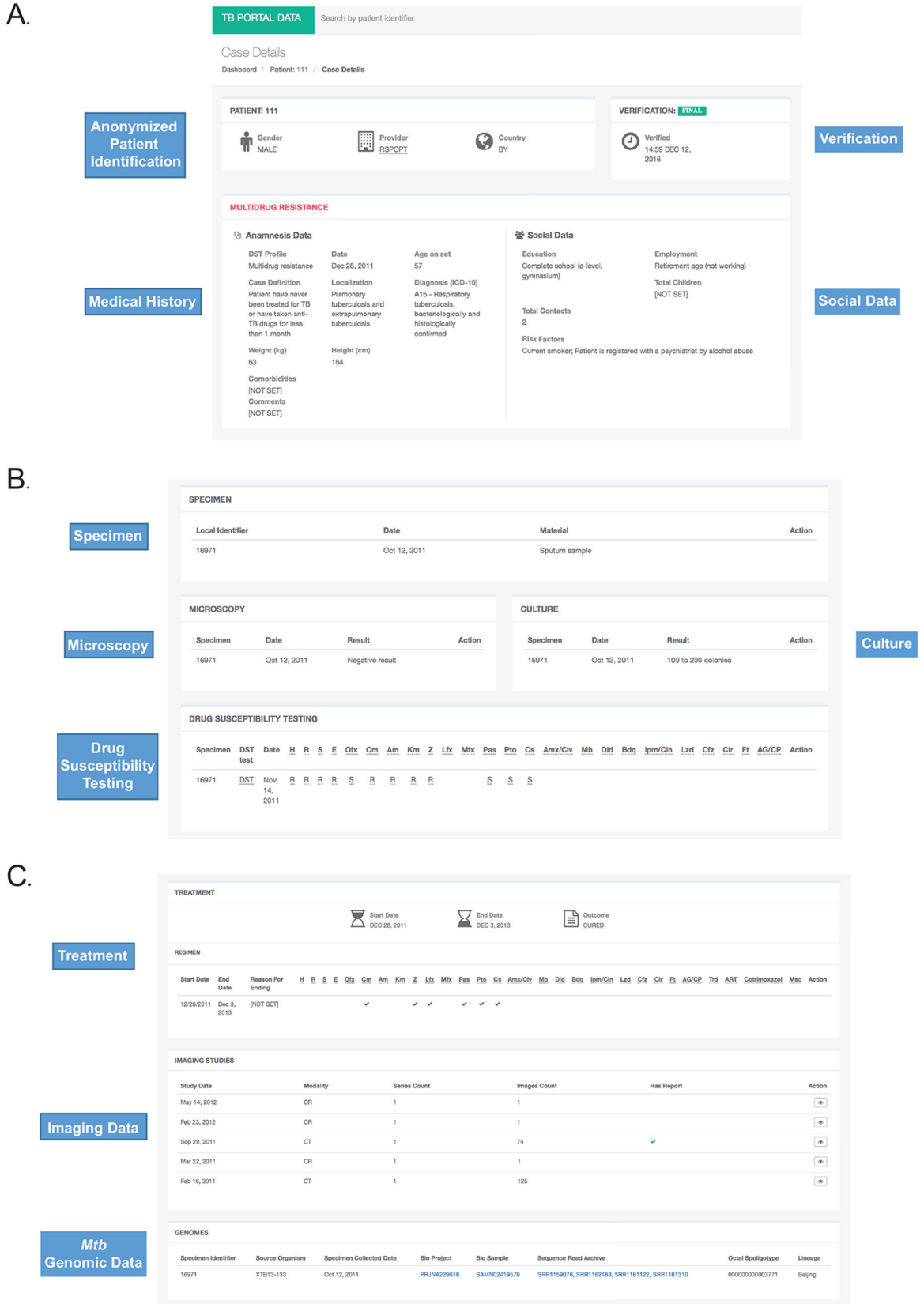

FIG 6 Patient case details. Included in each patient's case detail is a full profile containing data connected to that patient. This example patient case record includes anonymized patient information, verification records, medical history, and social data (A); specimen microscopy, culture, and drug susceptibility testing details (B); and treatment history, imaging data, and *M. tuberculosis* (*Mtb*) genomic information (C).

We carefully chose features or data points for each case in the TB Portals to concisely reflect the most important information about a patient diagnosed with MDR-TB (Table 1 and Fig. 6). Demographic details include standard measures of age, height, and weight, etc. Social data describe domicile and lifestyle. Clinical details include symptoms, comorbidities, treatment plans, and laboratory testing specifics, such as TB diagnostics used and results from smear microscopy, Xpert/RIF tests, line probe assays, culture on liquid and solid media, and phenotypic and genotypic drug sensitivity (susceptibility is the more commonly used term) testing (DST). Imaging data include original CXR radiographs (whether digitally obtained or on digitized film) and, in some cases, accompanying CT data sets. Genomics data sets are produced from *Mycobacterium tuberculosis* DNAs, deposited in the NCBI SRA, and linked to the associated TB Portal clinical specimens and cases.

The TB Portals Steering Committee plans for the set of clinical features and descriptors to be continuously updated, refined, and expanded. For example, medical imaging data will be annotated by qualified specialists using controlled vocabulary descriptors, and genomics descriptors will be added to the database as a result of ongoing sequencing projects. Currently, the genomic data fields contain simple links to sequence reads stored in the NCBI GenBank SRA, but we plan to expand these descriptors to indicate the presence of drug resistance mutations (variants such as SNPs and indels). As new descriptors and data types are added to the database, the TB Portals dictionary will be updated accordingly.

Imaging data. Clinical lung images, including X-ray and CT scans, are a distinct category of data associated with patient cases in the TB Portals. X-ray images are presented in frontal view only using Digital Imaging and Communications in Medicine (DICOM) file format. If unavailable in DICOM format, X-ray images are provided in Joint Photographic Experts Group (JPEG) file format at a 512- by 512-pixel resolution. CT images are presented in DICOM file format of various resolutions. All image files are manually and electronically reviewed to remove personally identifiable information and nonlung images. When available, information about an image, as reviewed and described by radiologists, is also included among patient case data.

Genomic data. We extract bacterial DNA from patient samples (sputum, biopsy specimens, surgical samples, bronchial lavage fluid, and ascitic fluid) and send it to one or more service facilities for whole-genome sequencing. Details of genomic sample processing were described previously by Wollenberg et al. (23). To date, most of the sequencing for the variants has been performed by the Broad Institute of MIT and Harvard but also by the Foundation for the Promotion of Health and Biomedical Research of Valencian Region, Valencia, Spain; by the Georgia NCDC; and by the National Reference Center for Mycobacteria, Borstel, Germany. The reads from all the samples processed by the Broad Institute and the National Reference Center for Mycobacteria, Germany, were submitted to the NCBI SRA. We are in the process of submitting the sequencing data from the other facilities to the NCBI SRA. We aligned the reads against the H37Rv reference genome (GenBank accession number CP003248.2) using BWA (24), called variants using Pilon v1.5 (25), and predicted the effects of these variants using either vcf annotator (https://sourceforge.net/projects/vcfannotator/) or SnpEff (http://snpeff.sourceforge.net/index.html). We then identified genomic polymorphisms that are known to confer drug resistance in each strain and that are described by the ReSeqTB database (https://platform.reseqtb.org/) as having either "large and often conclusive evidence" or "moderate evidence of drug resistance." We also performed quality assurance (QA)/quality control (QC) to check for sufficient coverage and possible contamination. To better understand the population structure of our samples, we determined lineage designations based on spoligotypes (26). Spoligotypes were determined and lineages assignments were made by using lorikeet software (10) and the TB-Lineage webpage (http://tbinsight.cs.rpi.edu/run_tb_lineage.html).

Open-access data. The TB Portals manifest the FAIR principles of data findability, accessibility, interoperability, and reusability (27). The mission of the TB Portals program revolves around the unification and distribution of data. The data are deidentified, properly curated, and validated to become a trusted accessible resource. Systems integration and interoperability enable data identification, input, display, and reporting throughout the technical ecosystem of the TB Portals. These open-access data enable the reuse of patient samples and extracted data for hypothesis generation and analysis yet to be identified. The display and reporting of these data are described in Results as well as in a future report on the TB Portals Data Exploration Portal (DEPOT).

ACKNOWLEDGMENTS

We acknowledge Andrea Bonhomme for technical writing support during the creation of the manuscript; the Broad Institute of MIT and Harvard for sequencing and analysis of the *M. tuberculosis* DNAs; the International Science and Technology Center for their support in establishing the TB Portal agreement with Georgia; CRDF Global for their support in establishing the TB Portal agreements with Azerbaijan, Belarus, Moldova, and Romania; and the NIAID Division of Microbiology and Infectious Diseases for their support with the *M. tuberculosis* sequencing efforts.

This research was supported in part by the Office of Science Management and Operations (OSMO) of the NIAID.

REFERENCES

1. Kim JY, Shakow A, Castro A, Vande C, Farmer P. 2003. Tuberculosis control. Global public goods for health: health economic and public health perspectives, p 54–72. Oxford University Press, New York, NY.
2. World Health Organization. 2016. Global tuberculosis report. World Health Organization, Geneva, Switzerland.
3. World Health Organization. 2006. The Stop TB Strategy. World Health Organization, Geneva, Switzerland. http://www.who.int/tb/strategy/stop_tb_strategy/en/. Accessed 23 June 2017.
4. World Health Organization. 2015. The End TB Strategy. World Health Organization, Geneva, Switzerland. http://www.who.int/tb/strategy/end-tb/en/. Accessed 23 June 2017.
5. Sharma A, Hill A, Kurbatova E, van der Walt M, Kvasnovsky C, Tupasi TE, Caoili JC, Gler MT, Volchenkov GV, Kazennyy BY, Demikhova OV, Bayona J, Contreras C, Yagui M, Leimane V, Cho SN, Kim HJ, Kliiman K, Akksilp S, Jou R, Ershova J, Dalton T, Cegielski P, Global Preserving Effective TB Treatment Study Investigators. 9 May 2017. Estimating the future burden of multidrug-resistant and extensively drug-resistant tuberculosis in India, the Philippines, Russia, and South Africa: a mathematical modelling study. Lancet Infect Dis https://doi.org/10.1016/S1473-3099(17)30247-5.
6. World Health Organization. 2016. Multidrug-resistant tuberculosis (MDR-TB) update. World Health Organization, Geneva, Switzerland. http://www.who.int/tb/challenges/mdr/mdr_tb_factsheet.pdf.
7. World Health Organization. 2017. Tuberculosis. World Health Organization, Geneva, Switzerland. http://www.who.int/mediacentre/factsheets/fs104/en/.
8. Yerlikaya S, Broger T, MacLean E, Pai M, Denkinger CM. 2017. A tuberculosis biomarker database: the key to novel TB diagnostics. Int J Infect Dis 56:253–257. https://doi.org/10.1016/j.ijid.2017.01.025.
9. World Health Organization. 2014. Definitions and reporting framework for tuberculosis—2013 revision. World Health Organization, Geneva, Switzerland.
10. Cohen KA, Abeel T, Manson McGuire A, Desjardins CA, Munsamy V, Shea TP, Walker BJ, Bantubani N, Almeida DV, Alvarado L, Chapman SB, Mvelase NR, Duffy EY, Fitzgerald MG, Govender P, Gujja S, Hamilton S, Howarth C, Larimer JD, Maharaj K, Pearson MD, Priest ME, Zeng Q, Padayatchi N, Grosset J, Young SK, Wortman J, Mlisana KP, O'Donnell MR, Birren BW, Bishai WR, Pym AS, Earl AM. 2015. Evolution of extensively drug-resistant tuberculosis over four decades: whole genome sequencing and dating analysis of Mycobacterium tuberculosis isolates from KwaZulu-Natal. PLoS Med 12:e1001880. https://doi.org/10.1371/journal.pmed.1001880.
11. Desjardins CA, Cohen KA, Munsamy V, Abeel T, Maharaj K, Walker BJ, Shea TP, Almeida DV, Manson AL, Salazar A, Padayatchi N, O'Donnell MR, Mlisana KP, Wortman J, Birren BW, Grosset J, Earl AM, Pym AS. 2016. Genomic and functional analyses of Mycobacterium tuberculosis strains implicate ald in D-cycloserine resistance. Nat Genet 48:544–551. https://doi.org/10.1038/ng.3548.
12. Zhang H, Li D, Zhao L, Fleming J, Lin N, Wang T, Liu Z, Li C, Galwey N, Deng J, Zhou Y, Zhu Y, Gao Y, Wang T, Wang S, Huang Y, Wang M, Zhong Q, Zhou L, Chen T, Zhou J, Yang R, Zhu G, Hang H, Zhang J, Li F, Wan K, Wang J, Zhang XE, Bi L. 2013. Genome sequencing of 161 Mycobacterium tuberculosis isolates from China identifies genes and intergenic regions associated with drug resistance. Nat Genet 45:1255–1260. https://doi.org/10.1038/ng.2735.
13. Farhat MR, Shapiro BJ, Kieser KJ, Sultana R, Jacobson KR, Victor TC, Warren RM, Streicher EM, Calver A, Sloutsky A, Kaur D, Posey JE, Plikaytis B, Oggioni MR, Gardy JL, Johnston JC, Rodrigues M, Tang PK, Kato-Maeda M, Borowsky ML, Muddukrishna B, Kreiswirth BN, Kurepina N, Galagan J, Gagneux S, Birren B, Rubin EJ, Lander ES, Sabeti PC, Murray M. 2013. Genomic analysis identifies targets of convergent positive selection in drug-resistant Mycobacterium tuberculosis. Nat Genet 45:1183–1189. https://doi.org/10.1038/ng.2747.
14. Casali N, Nikolayevskyy V, Balabanova Y, Harris SR, Ignatyeva O, Kontsevaya I, Corander J, Bryant J, Parkhill J, Nejentsev S, Horstmann RD, Brown T, Drobniewski F. 2014. Evolution and transmission of drug-resistant tuberculosis in a Russian population. Nat Genet 46:279–286. https://doi.org/10.1038/ng.2878.
15. Ioerger TR, Koo S, No EG, Chen X, Larsen MH, Jacobs WR, Jr, Pillay M, Sturm AW, Sacchettini JC. 2009. Genome analysis of multi- and extensively-drug-resistant tuberculosis from KwaZulu-Natal, South Africa. PLoS One 4:e7778. https://doi.org/10.1371/journal.pone.0007778.
16. Manson AL, Cohen KA, Abeel T, Desjardins CA, Armstrong DT, Barry CE, III, Brand J, TBResist Global Genome Consortium, Chapman SB, Cho SN, Gabrielian A, Gomez J, Jodals AM, Joloba M, Jureen P, Lee JS, Malinga L, Maiga M, Nordenberg D, Noroc E, Romancenco E, Salazar A, Ssengooba W, Velayati AA, Winglee K, Zalutskaya A, Via LE, Cassell GH, Dorman SE, Ellner J, Farnia P, Galagan JE, Rosenthal A, Crudu V, Homorodean D, Hsueh PR, Narayanan S, Pym AS, Skrahina A, Swaminathan S, Van der Walt M, Alland D, Bishai WR, Cohen T, Hoffner S, Birren BW, Earl AM. 2017. Genomic analysis of globally diverse Mycobacterium tuberculosis strains provides insights into the emergence and spread of multidrug resistance. Nat Genet 49:395–402. https://doi.org/10.1038/ng.3767.
17. Jaeger S, Karargyris A, Candemir S, Folio L, Siegelman J, Callaghan F, Xue Z, Palaniappan K, Singh RK, Antani S, Thoma G, Wang Y-X, Lu P-X, McDonald CJ. 2014. Automatic tuberculosis screening using chest radiographs. IEEE Trans Med Imaging 33:233–245. https://doi.org/10.1109/TMI.2013.2284099.
18. Kovalev V, Liauchuk V, Kalinovsky A, Rosenthal A, Gabrielian A, Skrahina A, Astrauko A, Tarasau A. 2015. CARS 2015—Computer Assisted Radiology and Surgery Proceedings of the 29th International Congress and Exhibition Barcelona, Spain, June 24-27, 2015: utilizing radiological images for predicting drug resistance of lung tuberculosis. Int J Comput Assist Radiol Surg 10(Suppl 1):1–312. https://doi.org/10.1007/s11548-015-1213-2.
19. Kovalev V, Liauchuk V, Safonau I, Astrauko A, Skrahina A, Tarasau A. 2013. Computed tomography: is there any correlation between the drug resistance and structural features of radiological images of lung tuberculosis patients. Int J Comput Assist Radiol Surg 8:S18–S20. https://doi.org/10.1007/s11548-013-0845-3.
20. Chen RY, Dodd LE, Lee M, Paripati P, Hammoud DA, Mountz JM, Jeon D, Zia N, Zahiri H, Coleman MT, Carroll MW, Lee JD, Jeong YJ, Herscovitch P, Lahouar S, Tartakovsky M, Rosenthal A, Somaiyya S, Lee S, Goldfeder LC, Cai Y, Via LE, Park SK, Cho SN, Barry CE, III. 2014. PET/CT imaging correlates with treatment outcome in patients with multidrug-resistant tuberculosis. Sci Transl Med 6:265ra166. https://doi.org/10.1126/scitranslmed.3009501.
21. World Health Organization. 2016. Use of high burden country lists for TB by WHO in the post-2015 era: summary. World Health Organization, Geneva, Switzerland.
22. Detjen AK, DiNardo AR, Leyden J, Steingart KR, Menzies D, Schiller I, Dendukuri N, Mandalakas AM. 2015. Xpert MTB/RIF assay for the diagnosis of pulmonary tuberculosis in children: a systematic review and meta-analysis. Lancet Respir Med 3:451–461. https://doi.org/10.1016/S2213-2600(15)00095-8.
23. Wollenberg KR, Desjardins CA, Zalutskaya A, Slodovnikova V, Oler AJ, Quinones M, Abeel T, Chapman SP, Tartakovsky M, Gabrielian A, Hoffner S, Skrahin A, Birren BW, Rosenthal A, Skrahina A, Earl AM. 2017. Whole-genome sequencing of Mycobacterium tuberculosis provides insight into the evolution and genetic composition of drug-resistant tuberculosis in Belarus. J Clin Microbiol 55:457–469. https://doi.org/10.1128/JCM.02116-16.
24. Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595. https://doi.org/10.1093/bioinformatics/btp698.
25. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
26. Vitol I, Driscoll J, Kreiswirth B, Kurepina N, Bennett KP. 2006. Identifying Mycobacterium tuberculosis complex strain families using spoligotypes. Infect Genet Evol 6:491–504. https://doi.org/10.1016/j.meegid.2006.03.003.
27. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJ, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PA, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, et al. 2016. The FAIR guiding principles for scientific data management and stewardship. Sci Data 3:160018. https://doi.org/10.1038/sdata.2016.18.